Communications for Statistical Applications and Methods
2013, Vol. 20, No. 1, 1–5

DOI: http://dx.doi.org/10.5351/CSAM.2013.20.1.001

Bootstrap aggregated classification for sparse functional data

Hyunsung Kim[a], Yaeji Lim[1,b]

[a]Department of Statistics, Chung-Ang University, Korea
[b]Department of Applied Statistics, Chung-Ang University, Korea

Abstract

Abstract

Keywords: functional principal component analysis, bootstrap aggregating, classification

1. Introduction

Prior studies

Description of sections

2. Preliminaries

2.1. Functional principal compnent analysis

Functional data analysis (FDA) is a kind of statistics that analyzes the curves or functions rather than single points. In FDA, data (or curve) is defined on the infinite dimension, so dimensionality reduction becomes a key issue. One of the popular dimension reduction method in FDA is functional principal component analysis (FPCA) which finds directions of the variation and exploits by data-driven basis called functional principal component (FPC) scores.

Let $X(t)$, defined on finite $\mathcal{T}$, is a square integrable random process which means $X(\cdot) \in L_2(\mathcal{T})$ with mean function $\mu(t) = E(X(t))$ and covariance function $G(s,t) = \text{cov}(X(s), X(t))$ for $s, t \in \mathcal{T}$. By Mercer's theorem, the covariance function can be represented by $G(s,t) = \sum_{k=1}^{\infty} \lambda_k \phi_k(s)\phi_k(t)$ where $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ are non-negative eigenvalues satisfying $\sum_{k=1}^{\infty} \lambda_k < \infty$ and ϕ_k's are the corresponding orthonormal eigenfunctions. By the Karhunen-Loève expansion, the ith random curve can be represented as

$$X_i(t) = \mu(t) + \sum_{k=1}^{\infty} \xi_{ik}\phi_k(t), \qquad t \in \mathcal{T} \tag{2.1}$$

where $\xi_{ik} = \int_{\mathcal{T}} (X_i(t) - \mu(t))\phi_k(t)dt$ are uncorrelated variables with mean 0 and variance λ_k.

In the real world, functional data is observed with measurement errors. Then the ith curve with random noise is

$$\begin{aligned} U_i(t) &= X_i(t) + \epsilon_i(t) \\ &= \mu(t) + \sum_{k=1}^{\infty} \xi_{ik}\phi_k(t) + \epsilon_i(t), \qquad t \in \mathcal{T} \end{aligned} \tag{2.2}$$

[1] Corresponding author: Professor, Department of Applied Statistics, Chung-Ang University, 84 Heukseok-ro, Dongjak-gu, Seoul 06974, Korea. E-mail: yaeji.lim@gmail.com

where $\epsilon_i(t)$ are the uncorrelated measurement errors with mean 0 and variance σ^2.

PC selection? PVE, AIC, BIC, truncation number(size) K

2.2. Functional principal component analysis for sparse functional data

Let $t_{ij} \in \mathcal{T}$ is the jth time point observed in the ith curve X_i where $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, n_i$. Then, the ith curve for sparse functional data can be expressed as

$$U_i(t_{ij}) = \mu(t_{ij}) + \sum_{k=1}^{\infty} \xi_{ik}\phi_k(t_{ij}) + \epsilon_i(t_{ij}), \qquad t_{ij} \in \mathcal{T} \tag{2.3}$$

where $\epsilon_i(t_{ij})$ are the random noises.

In dense functional data which means that data is observed at regular grid, we can estimate the FPC scores by numerical integration directly. On the other hand, when each curve is observed at sparse or irregular time points, estimating the FPC scores by numerical integration is not appropriate since the estimate of ξ_{ik} would be biased. To overcome this problem, Yao *et al.* (2005) proposed principal component analysis through conditional expectation (PACE) method. Let $\mathbf{X}_i = (X_i(t_{i1}), \ldots, X_i(t_{in_i}))^T$, $\mathbf{U}_i = (U_i(t_{in_1}), \ldots, U_i(t_{in_i}))^T$, $\boldsymbol{\mu}_i = (\mu(t_{i1}), \ldots, \mu(t_{in_i}))^T$, $\boldsymbol{\phi}_{ik} = (\phi_k(t_{i1}), \ldots, \phi_k(t_{in_i}))^T$, and $\boldsymbol{\epsilon}_i = (\epsilon_i(t_{i1}), \ldots, \epsilon_i(t_{in_i}))^T$. The best linear unbiased prediction (BLUP) of ξ_{ik}, kth FPC scores of ith subject, by PACE is

$$\tilde{\xi}_{ik} = E[\xi_{ik}|\mathbf{U}_i] = \lambda_k \boldsymbol{\phi}_{ik}^T \boldsymbol{\Sigma}_{\mathbf{U}_i}^{-1}(\mathbf{U}_i - \boldsymbol{\mu}_i), \tag{2.4}$$

where $\boldsymbol{\Sigma}_{\mathbf{U}_i} = \mathrm{cov}(\mathbf{U}_i, \mathbf{U}_i) = \mathrm{cov}(\mathbf{X}_i, \mathbf{X}_i) + \sigma^2 \mathbf{I}_{n_i}$,

From the above, the PACE estimate of ξ_{ik} is obtained as follow

$$\hat{\xi}_{ik} = \widehat{E}[\xi_{ik}|\mathbf{U}_i] = \hat{\lambda}_k \hat{\boldsymbol{\phi}}_{ik}^T \widehat{\boldsymbol{\Sigma}}_{\mathbf{U}_i}^{-1}(\mathbf{U}_i - \hat{\boldsymbol{\mu}}_i). \tag{2.5}$$

where $\widehat{\boldsymbol{\Sigma}}_{\mathbf{U}_i} = \widehat{\mathrm{cov}}(\mathbf{U}_i, \mathbf{U}_i) = \widehat{\mathrm{cov}}(\mathbf{X}_i, \mathbf{X}_i) + \hat{\sigma}^2 \mathbf{I}_{n_i}$

How to apply FPCA for sparse data?

3. Method

3.1. Functional classification(Classification based on functional principal component scores)

Let $X_i(\cdot)$ is the ith curve, Y_i is its class response, $\beta(\cdot)$ is the weight function and ϵ_i is the ith random error. Then, the functional linear model (FLM) is

$$Y_i = \int_{\mathcal{T}} X_i(t)\beta(t)dt + \epsilon_i. \tag{3.1}$$

If we use the FPCA form that means $X_i(t) = \sum_{k=1}^{\infty} \xi_{ik}\phi_k(t_{ij})$ and $\beta(t) = \sum_{k=1}^{\infty} \beta_k \phi_k(t_{ij})$, then the FLM is represented by

$$\begin{aligned} Y_i &= \sum_{k=1}^{\infty} \beta_k \xi_{ik} + \epsilon_i \\ &\approx \sum_{k=1}^{K} \beta_k \xi_{ik} + \epsilon_i, \end{aligned} \tag{3.2}$$

where K is the truncated number on FPCA.

3.2. Bootstrap aggregating

Bootstrap aggregating (Bagging) is the ensemble method using bootstrap ideas proposed by Breiman (1996). It is known to improve performance of the classifier and reduce the variance by aggregating predictions from each bootstrap resample. Suppose there are K response classes with $C_1, \ldots, C_K$ and $\hat{f}^{(b)}(x)$ for $b = 1, \ldots, B$ be a classifier obtained by bth bootstrap resample. Then, the estimate of bagged classifier using the majority vote scheme is

$$\hat{y}_{\text{bag}} = \arg\max_j \#\{b | \hat{f}^{(b)}(x) = C_j\}. \tag{3.3}$$

What is bagging?

3.3. Bootstrap aggregated classifier with sparse FPCA

The proposed method is bootstrap aggregated classification with FPCA for sparse functional data. Suppose we have n curves with K different class label denoted by $\mathcal{D} = \{(\mathbf{U}_1, y_1), \ldots, (\mathbf{U}_n, y_n)\}$. Denote $\mathcal{D}^{(b)} = \{(\mathbf{U}_1^{(b)}, y_1^{(b)}), \ldots, (\mathbf{U}_n^{(b)}, y_n^{(b)})\}$ is a bth bootstrap resample.

The summary of procedure is below algorithm.

Algorithm 1: Bagged classifier with sparse FPCA

```
1. For $b = 1, \ldots, B$, repeat
    (a) Generate a bootstrap resample $\mathcal{D}^{(b)}$ from the data $\mathcal{D}$.
    (b) Apply functional principal component analysis for $\mathcal{D}^{(b)}$.
    (c) Estimate the FPC scores by PACE.
    (d) Select $K$, the number of FPCs, such that PVE $\geq 0.99$.
    (e) Construct classifier on $K$ FPC scores.
2. Given a new curve $\mathbf{U}^*$, for $b = 1, \ldots, B$, repeat
    (a) Estimate the FPC scores by PACE using FPC function from $\mathcal{D}_b$.
    (b) Obtain the prediction $\hat{y}^{(b)}$ from above $b$th classifiers.
3. From $\hat{y}^{(1)}, \ldots, \hat{y}^{(B)}$, obtain the final prediction $\hat{y}_{bag}$ aggreated by majority vote.
```

algorithm flow

algorithm image

What is the proposed method?

4. Simulation studies

Data generating

Result

5. Real data analysis

5.1. Berkely growth data

We applied the proposed method to the berkely growth data presented in Tuddenham & Snyder (1954). The dataset was measured height for 93 individuals, 54 girls and 39 boys. There are 31 observations from ages 1 to 18 for each curve. These dense curves with 54 girls and 39 boys are plotted in Figure 1.

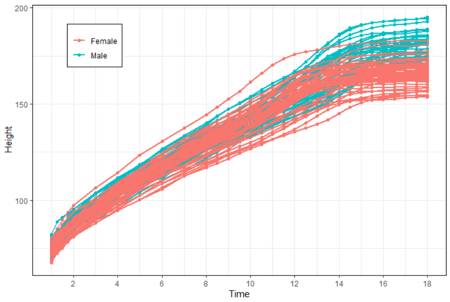

Figure 1: *The berkely growth data of 93 individuals.*

Table 1: The average classification error with standard error in percentage from 100 Monte Carlo repetitions for berkerly growth data

	Logistic	SVM	SVM			Naive
Method	Regression	(Linear)	(Gaussian)	LDA	QDA	Bayes
Single	7.3 (4.80)	5.3 (3.20)	5.7 (4.03)	5.8 (3.34)	5.6 (3.35)	5.6 (3.90)
Majority vote	5.9 (4.12)	4.9 (3.19)	5.3 (3.51)	5.4 (3.24)	4.9 (3.57)	5.5 (3.96)
OOB weight	5.9 (4.12)	5.0 (3.22)	5.4 (3.62)	5.4 (3.27)	4.9 (3.54)	5.5 (3.96)

Data description

Result

5.2. Spinal bone mineral density data

We applied the proposed method to the spinal bone mineral density data presented in Bachrach *et al.* (1999). The dataset has the spinal bone mineral density for 280 individuals, 153 females and 127 males measured at sparse and irregualar time points. There are 2 4 observations for each curve. The dataset has also ethnicity information such that Asian, Black, Hispanic and White. These sparse curves with 153 females and 127 males are plotted in Figure 2.

We try to classify gender for proposed method. First, we split the sparse spinal bone mineral density data to training and test set. We generate 100 bootstrap resamples from training set. We applied FPCA with PACE for each bootstrap resample to obtain the FPC scores. We construct 100 classifiers on FPC scores from each bootstrap resample. From test set, 100 bootstrap predictions are obtained for each curve. Then we get the final prediction by majority vote from 100 bootstrap predictions. We repeat 100 times with different seeds for different split of training and test set.

Data description

Result

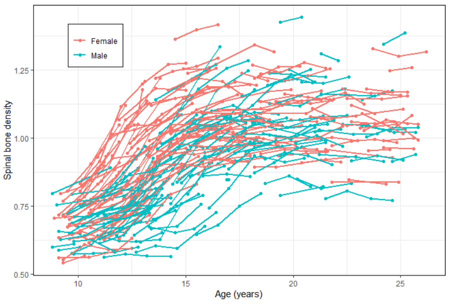

Figure 2: *The spinal bone mineral density of 280 individuals.*

Table 2: The average classification error with standard error in percentage from 100 Monte Carlo repetitions for spinal bone mineral density data

Method	Logistic Regression	SVM (Linear)	SVM (Gaussian)	LDA	QDA	Naive Bayes
Single	31.3 (4.30)	32.0 (4.27)	33.2 (4.71)	31.4 (4.44)	33.3 (4.10)	32.3 (4.33)
Majority vote	30.2 (3.72)	30.8 (4.18)	31.2 (3.88)	30.4 (3.77)	31.6 (3.78)	30.9 (3.83)
OOB weight	30.3 (3.71)	30.8 (4.07)	31.4 (3.81)	30.5 (3.82)	31.8 (3.71)	30.9 (3.86)

6. Conclusion and discussion

conclusion and result

Acknowledgement

This research was supported by the Chung-Ang University Graduate Research Scholarship in 2019.

References

Bachrach, L. K., Hastie, T., Wang, M. C., Narasimhan, B., & Marcus, R. (1999). Bone mineral acquisition in healthy Asian, Hispanic, black, and Caucasian youth: a longitudinal study. *The journal of clinical endocrinology & metabolism*, **84**, 4702-4712.

Breiman, L. (1996). Bagging predictors. *Machine learning*, **24**, 123-140.

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning.*, Springer series in statistics, New York.

James, G. M., Hastie, T. J., & Sugar, C. A. (2000). Principal component models for sparse functional data. *Biometrika*, **87**, 587-602.

James, G. M., & Hastie, T. J. (2001). Functional linear discriminant analysis for irregularly sampled curves. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **63**, 533-550.

Ramsay, J. O., & Silverman, B. W. *Functional Data Analysis*, 2nd ed., Springer Series in Statistics, New York.

Tuddenham, R. D., & Snyder, M. M. (1954). Physical growth of California boys and girls from birth to eighteen years. *University of California publications in child development*, **1**, 183-364.

Yao, F., Müller, H. G., & Wang, J. L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, **100**, 577-590.

Received 00, 0000; Revised 00, 0000; Accepted 00, 0000